7-2011

# Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats

Franco G. Asoro
*Iowa State University*

Mark A. Newell
*Iowa State University*, markandew.newell@gmail.com

William D. Beavis
*Iowa State University*, wdbeavis@iastate.edu

Marvin Paul Scott
*United States Department of Agriculture*, pscott@iastate.edu

Jean-Luc Jannink
*United States Department of Agriculture*

# Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats

**Abstract**
Genomic selection (GS) is a method to estimate the breeding values of individuals by using markers throughout the genome. We evaluated the accuracies of GS using data from five traits on 446 oat (*Avena sativa* L.) lines genotyped with 1005 Diversity Array Technology (DArT) markers and two GS methods (ridge regression–best linear unbiased prediction [RR-BLUP] and BayesCπ) under various training designs. Our objectives were to (i) determine accuracy under increasing marker density and training population size, (ii) assess accuracies when data is divided over time, and (iii) examine accuracy in the presence of population structure. Accuracy increased as the number of markers and training size become larger. Including older lines in the training population increased or maintained accuracy, indicating that older generations retained information useful for predicting validation populations. The presence of population structure affected accuracy: when training and validation subpopulations were closely related accuracy was greater than when they were distantly related, implying that linkage disequilibrium (LD) relationships changed across subpopulations. Across many scenarios involving large training populations, the accuracy of BayesCπ and RR-BLUP did not differ. This empirical study provided evidence regarding the application of GS to hasten the delivery of cultivars through the use of inexpensive and abundant molecular markers available to the public sector.

**Disciplines**
Agronomy and Crop Sciences | Plant Breeding and Genetics

**Comments**
This article is from *The Plant Genome* 4 (2011): 132, doi: 10.3835/plantgenome2011.02.0007.

Published July, 2011

ORIGINAL RESEARCH

# Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats

Franco G. Asoro, Mark A. Newell, William D. Beavis, M. Paul Scott, and Jean-Luc Jannink*

## Abstract

Genomic selection (GS) is a method to estimate the breeding values of individuals by using markers throughout the genome. We evaluated the accuracies of GS using data from five traits on 446 oat (*Avena sativa* L.) lines genotyped with 1005 Diversity Array Technology (DArT) markers and two GS methods (ridge regression–best linear unbiased prediction [RR-BLUP] and BayesC$\pi$) under various training designs. Our objectives were to (i) determine accuracy under increasing marker density and training population size, (ii) assess accuracies when data is divided over time, and (iii) examine accuracy in the presence of population structure. Accuracy increased as the number of markers and training size become larger. Including older lines in the training population increased or maintained accuracy, indicating that older generations retained information useful for predicting validation populations. The presence of population structure affected accuracy: when training and validation subpopulations were closely related accuracy was greater than when they were distantly related, implying that linkage disequilibrium (LD) relationships changed across subpopulations. Across many scenarios involving large training populations, the accuracy of BayesC$\pi$ and RR-BLUP did not differ. This empirical study provided evidence regarding the application of GS to hasten the delivery of cultivars through the use of inexpensive and abundant molecular markers available to the public sector.

Published in The Plant Genome 4:132–144. Published 21 June 2011.
doi: 10.3835/plantgenome2011.02.0007

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication



THE DECREASING COST of high-density molecular markers allows saturation of crop genomes with genetic markers and offers an approach to predict genetic merit. These markers can help capture the effects of many quantitative trait loci (QTL) controlling polygenic traits regardless of location of the QTL in the genome by using linkage disequilibrium (LD), the nonrandom association of alleles at different loci (Falconer and Mackay, 1996). Meuwissen et al. (2001) proposed genomic selection (GS) based on prediction of the genetic value of individuals or the genomic estimated breeding values (GEBV) from high-density markers positioned throughout the genome. Because GS includes all markers, major and polygenic effects can be captured, potentially explaining more genetic variance (Solberg et al., 2008). Therefore, the objective of GS is to predict the breeding value of each individual instead of identifying QTL for use in a traditional marker-assisted selection (MAS) program.

Selection methods can be evaluated by measuring accuracy, a major component of the response to selection equation, $R = ir\sigma_A$, in which $R$ is the response, $i$ is the selection intensity, $r$ is the accuracy, and $\sigma_A$ is the additive genetic standard deviation (Falconer and Mackay, 1996). As a general term in statistics, accuracy is the degree of similarity

F.G. Asoro, M.A. Newell, and W.D. Beavis, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; M.P. Scott, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Dep. of Agronomy, Ames, IA 50011; J.-L. Jannink, USDA-ARS, R.W. Holley Center for Agriculture and Health, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14853. Received 25 Feb. 2011. *Corresponding author (jeanluc.jannink@ars.usda.gov).

**Abbreviations:** ARS, Agricultural Research Service; BLUP, best linear unbiased prediction; C, cluster; GBLUP, genomic relationship BLUP; GEBV, genomic estimated breeding values; GS, genomic selection; GWAS, genome-wide association studies; LD, linkage disequilibrium; PC, principal component; QTL, quantitative trait loci; QUON, Quaker Uniform Oat Nursery; RR-BLUP, ridge regression–best linear unbiased prediction; TBV, true breeding value; UOPN, Uniform Oat Performance Nursery.

between the true value and the estimated value (Taylor, 1997). In crop selection programs, accuracy is defined as the correlation between the phenotype of the selected lines, that is, selection units, and the phenotype transmitted to the progeny of the selected lines, that is, response units (Holland et al., 2003). If the response population is composed of progeny of selected individuals, then accuracy is the correlation between the selection criterion and the true breeding value (TBV; Falconer and Mackay, 1996), since breeding values are by definition the mean of the progeny of individuals. If the selection criterion is the individual's phenotypic performance, $r$ is equal to the square root of the heritability (Falconer and Mackay, 1996). In empirical cross-validation studies of GS, the TBV is unknown and, to compute accuracy, the TBV must be replaced by the traditional pedigree-based best linear unbiased prediction (BLUP) values, the least squares means from phenotypic evaluation, or some other appropriate phenotypic measurement (Garrick et al., 2009). The relationship between TBV and GEBV in the context of response to selection is explained in detail by Dekkers (2007).

Genomic selection in plant breeding has been studied in different types of populations. For example, GS has been used in narrow-based biparental populations (Lorenzana and Bernardo, 2009) and in broad-based populations such as multilines of barley (*Hordeum vulgare* L.), wheat (*Triticum aestivum* L.), and maize (*Zea mays* L.) (Zhong et al., 2009; Heffner et al., 2011; de los Campos et al., 2009; Crossa et al., 2010). Regardless of the type of population used, the basic steps for implementation of GS can be summarized in four steps: (i) designing training populations with complete phenotypic and genotypic data, (ii) estimating marker effects in the training population, (iii) calculating GEBV of new breeding lines with genotype data, and (iv) selection (Heffner et al., 2009; Jannink et al., 2010). Different methods exist to implement GS given the complexity of estimating marker effects to predict GEBV. These methods include ridge regression–best linear unbiased prediction (RR-BLUP) and equivalent methods based on genomic relationships (e.g., "realized A matrix BLUP" [RA-BLUP] in Zhong et al. [2009] or "genomic relationship BLUP" [GBLUP] in Habier et al. [2007]) and Bayesian-based methods such as BayesA, BayesB, BayesCπ, and BayesLASSO (Meuwissen et al., 2001; Kizilkaya et al., 2010; de los Campos et al., 2009). One important difference between RR-BLUP and the Bayesian methods is the prior distribution for the variance of marker effects: the former assigns equal variance to all markers while the latter allows unequal variances for markers. In numerous simulations and a few empirical studies of GS in both plants and animals, it has been shown that factors affecting accuracy include the genetic architecture of the trait, LD, genetic relationships between training and validation populations, marker density, training population size, and heritability (Hayes et al., 2009; Zhong et al., 2009, Luan et al., 2009, Daetwyler et al., 2010b, de Roos et al., 2009). In an empirical crossvalidation study of biparental plant populations, Lorenzana and Bernardo (2009) demonstrated that accuracy increases with training population size. It was also shown that increasing the number of markers generally resulted in increased accuracy, but the increase was large only at low marker densities. For instance, in their study of grain protein content in the 'Steptoe' × 'Morex' doubled haploid barley population, there was a clear increase in accuracy when changing from 64 to 128 markers; however, accuracy did not change from 128 to 223 markers.

Population structure or differing levels of relatedness of individuals in a population can have an impact on genome-wide studies. It has been demonstrated that accounting for population structure avoids spurious associations in genome-wide association studies (GWAS) (Yu et al., 2006). In GS, while population structure is still relevant, the focus shifts to maintaining the accuracy across different subpopulations or germplasm groupings (Lorenz et al., 2011). In the simulation study of Toosi et al. (2010), accuracy was high when the training population and validation population belonged to the same breed of animals, but they also showed that there was no substantial decrease of accuracy when a multibreed training population was used to estimate marker effects. In the empirical study of Hayes et al. (2009), GEBV predictions were more accurate within breed (e.g., Jersey to Jersey) than across breeds (e.g., Jersey to Holstein). However, when they used a multibreed training population (Jersey and Holstein) to predict purebred individuals (Jersey or Holstein), they found comparable accuracies as for the within-breed predictions. Developing a multisubpopulation training population is another way to increase training size and this approach may be important if subpopulations are small (de Roos et al., 2009). Although these studies suggest the importance of genetic relationships of the training and validation population, more importantly they indicate that in the presence of population structure, LD should be consistent across subpopulations to maintain accuracy. This means that allelic effects estimated in one population should be predictive in another population (Lorenz et al., 2011). Such consistency of LD, however, requires higher marker densities (Meuwissen, 2009, Hamblin et al., 2010, Newell et al., 2010), and it is not clear if such densities are available for oat (*Avena sativa* L.).

Currently there are few empirical studies of GS in crops. Thus, while simulations have shown that these methods have great potential, we do not know how well they will work in practice. Studies in several species and populations will be necessary to gain a general appreciation for investments in the marker density and training population size. As a case study, we evaluated the accuracies of GS for five traits in oats (grain β-glucan content, yield, heading date, groat percentage, and plant height) from a public cooperative testing network in North America. The lines tested in the trials represent the breadth of alleles present in elite oat breeding populations; thus, they are a good sample for cross validation with potential impact in applied breeding programs. In this population, we assess the impact of marker density and training population size. This population is also structured so that we can present the first results in crops on the impact of structure on GS accuracy. Finally, RR-BLUP

and BayesCπ have only been compared in simulation studies (Jannink, 2010) and here we provide a comparison using empirical data.

## Materials and Methods

### Phenotypic Data Analysis

The majority of phenotypic data for β-glucan percentage, yield, heading date, groat percentage, and plant height of oat breeding lines and cultivars included in this study came from the Uniform Oat Performance Nursery (UOPN) and the Quaker Uniform Oat Nursery (QUON) from 1994 to 2007 (Matthews, 2011). The UOPN is a cooperative testing network for oats among different U.S. State Agricultural Experiment Stations and the USDA-Agricultural Research Service (ARS). The QUON is a cooperative testing network for oats among northern U.S. State Agricultural Experimental Stations, USDA-ARS, and public breeding institutions in Canada. Data for β-glucan percentage was also included from research conducted by Chernyshova et al. (2007) and Colleoni-Sirghie et al. (2004). In total, there were 446 oat lines with β-glucan data and 421 lines with data for the four remaining traits. Data came from 129 environments (combination of years and locations) for β-glucan, 328 for days to heading, 278 for groat percentage, 354 for plant height, and 388 for yield. Since not all of the lines were tested in the same environments, statistical analysis of this highly unbalanced data was conducted using PROC Mixed in SAS (SAS Institute, 2008), with environments considered fixed effects and oat lines as independently and identically distributed random effects. In this case, environments were considered as fixed effects to remove the effects of the mean of sets of environments on the genotypic performance due to the fact that some lines were tested in few locations or some years only. As such, oat lines were treated as random effects as they are considered a sample of all possible oat genotypes. The BLUP for each line was used as its observed phenotypic value and denoted y*.

### Marker Data, Relationship Matrix, and Population Structure

Lines were planted in the Iowa State University Agronomy greenhouse (Ames, IA) in Spring 2008, leaf samples were collected for each entry, and DNA was extracted according to the recommended protocol for Diversity Array Technology (DArT) markers (Diversity Arrays Technology, 2011). Deoxyribonucleic acid samples were then sent to Diversity Arrays Technology (Yarralumla, Australia) for genotyping. Diversity Array Technology markers are a dominant marker system; thus, for each of the 1295 markers, oat lines were scored for presence (1) or absence (0) of hybridization signal using a microarray platform (Tinker et al., 2009).

To eliminate redundant markers, sets of markers in perfect linkage disequilibrium (i.e., the squared correlation between marker scores was equal to 1) were identified. The marker with the lowest number of missing data points in each set was used in this study, resulting in 1005 markers.

To compute the marker-based relationship matrix, genotypic data points scored as absent (0) were recoded as –1, resulting in a data matrix of –1s and 1s. For each marker, missing values were replaced by the mean for that marker. The recoded marker matrix, **M,** was then used to compute the **MM′** matrix, which was divided by 1005, scaling the relationship values from 0 to 1 in which the minimum value was 0.01 and the maximum value was 1.00. To account for population structure, principal component analysis (PCA) was applied to the relationship matrix. The first five principal components (PCs), which explained about 76% of variation in the marker data, were chosen based on the scree plot (Cattell, 1966). The corresponding five eigenvectors were used as fixed population structure covariates. Principal components have been used as another way to correct for population structure in GWAS and LD studies (Price et al., 2006; Stich et al., 2008; Newell et al., 2010).

### Methods of Genomic Selection and Prediction of Genomic Estimated Breeding Values

The general model used was: $y^* = \mu + \mathbf{Q}\nu + \mathbf{M}\alpha + e$, in which $y^*$ is the observed phenotypic value, μ is the intercept, $\mathbf{Q}\nu$ is a fixed effects term where **Q** is a matrix of the first five PC eigenvectors, and ν is a vector of regression coefficients relating the first five PCs to the observed phenotype. The $\mathbf{Q}\nu$ term was excluded in the cluster-based training design (see below) because the clustering itself accounted for population structure. The $\mathbf{M}\alpha$ is a random effects term where **M** is the marker matrix and α is a vector of estimated marker effects.

Marker effects for RR-BLUP were simultaneously estimated and drawn from a normal distribution with equal variance, $N(0, \sigma_a^2)$ (Meuwissen et al., 2001). This method was implemented in the computer software R (R Development Core Team, 2009) using the emma package (Kang et al., 2008) and matrix algebra functions, in which the emma.MLE function was used to the estimate variance components $\sigma^2_{genetic}$ and $\sigma^2_{error}$ and the shrinkage parameter. The variance components and shrinkage parameter above were estimated in every sample of the training population. Finally, the shrinkage parameter computed above was incorporated in the mixed model equations to predict the marker effects.

For the BayesCπ method, described by R.L. Fernando (personal communication, 2010), markers are represented as random effects (α) and are normally distributed when included in the model but equal to 0 when not included in the model with prior probability π. In contrast to BayesB (Meuwissen et al., 2001), the π parameter is estimated from the data. Further, the marker variance for BayesCπ, $\sigma_a^2$, is assumed a priori to be distributed as a scaled inverse $\chi^2$ as explained in detail in Kizilkaya et al. (2010). A total of 1000 burn-in and 4000 saved iterations of Markov-chain Monte Carlo (MCMC) were used for BayesCπ in all designs. This method was implemented in R using code written by R.L. Fernando (personal communication, 2010).

Marker effects estimated from RR-BLUP and BayesCπ were used to predict the estimated genotypic values for the validation population. The GEBV prediction model was $GEBV = M\hat{\alpha}$, in which **M** is the marker matrix and $\hat{\alpha}$ is are the estimated marker effects.

## Design of Training and Validation Populations

To implement crossvalidation for accuracy of GEBV, the observed phenotypic values (y*) for all lines were divided into training and validation data sets using three different methods:

1. Random Lines and Markers. Training populations were selected at random with the restriction that descendants of any individual in the validation population were excluded (to the extent possible given pedigree records available). We implemented this restriction because training populations will rarely contain descendants of selection candidates in practice and because descendants contain information about the Mendelian sampling term entering the breeding value of an individual (Falconer and Mackay, 1996), whereas collateral relatives will not. Including descendants would therefore bias accuracies upward. Sets of 100, 200, and 300 lines were used as training populations while the remaining lines were used as validation populations with all 1005 markers retained. To determine the effect of marker density on accuracy, randomly selected sets of 300, 600, and 900 markers were used with a training population of 300 lines selected as describe above.
2. Testing Year-Grouping of Lines. Lines were grouped based on their first year of entry in the uniform nurseries. Years grouped as 1994 through 2003, 1998 through 2003, and 2001 through 2003 gave similar-sized training populations as for the randomly selected lines, resulting in 292, 220, and 106 for β-glucan and 282, 213, and 99 for all other traits, respectively. To remove the effect of unequal training population sizes across traits, a random sample of 90 lines from 2001 through 2003, 180 lines from 1998 through 2003, and 270 lines from 1994 through 2003 were chosen as the final training population for 100 replicates. These training populations confound changes in size with changes in age. They do, however, answer the practically important question of the utility of increasing the training population size by adding older (historical) lines to the training population. To avoid confounding of training population size on training population age, another two sets of training population from 1994 through 1998 and 1998 through 2000 with 90 randomly selected lines each were also developed for comparison to the training population from 2001 through 2003. For all of these designs, the validation population consisted of lines from the 2004 through 2007 yr grouping, which included 154 lines for β-glucan and 139 lines for the remaining traits.
3. Cluster-Based Grouping of Lines. For grouping the oat germplasm, the relationship matrix among the 446 lines was converted to a distance matrix by subtracting the values from one. Hierarchical clustering using Ward's linkage was applied to the distance matrix and implemented using the hclust function in the computer software R (R Development Core Team, 2009). Three clusters were chosen for two reasons: (i) to maximize the number of individuals in each cluster and (ii) the clustering produced two more related clusters and one less related cluster (Supplemental Fig. S1). The cluster dendrogram indicated that cluster 2 (C2) and cluster 3 (C3) are more highly related to each other than either is to cluster 1 (C1). The clusters C1, C2, and C3 consisted of 130, 179, and 137 lines, respectively, for β-glucan, and 128, 172, and 121 lines, respectively, for the other traits. A random sample of 120 lines from each cluster was used as the training population, while the other two clusters were used as validation populations. Additionally, to examine the effect of using combined clusters and training population size in accuracy, random samples of 60 and 120 lines were taken from each of two clusters and combined to serve as 120 and 240 line training populations while the remaining cluster was used for the validation population.

For each of these designs, results were based on the average from 100 random replicates of the training populations.

## Accuracy

Accuracy, calculated as the correlation of the observed (y*) and predicted breeding values (GEBV) in the validation sets was computed for each training design. Since population structure effects were in the model in the first two training designs, the accuracy was calculated to account for population structure effects in the y* vector by using the correlation (y* – **Q**ν, GEBV). This adjusted correlation will reflect the accuracy of GEBV excluding the variation due to population structure. The GEBV, with this adjustment, predicted within-subpopulation or within-cluster variation rather than all variation, which combined within- and between-subpopulation variation.

## Comparison of Accuracies

To compare how accuracy was affected by different GS methods, traits, and training population designs, ANOVA was conducted for each training population-validation population design with the following model:

$$r = \mu + trait + method + design + (trait \times method) + (trait \times design) + (method \times design) + error,$$

in which μ is the mean accuracy, the levels of *trait* are the five traits in this study, the levels of *method* are either

**Table 1. Analysis of variance (ANOVA) *p*-values for factors affecting the accuracies when designing training population with different numbers of markers, number of lines, lines sampled deeper in time, and line of different ages.**

| Source of variation | df | Marker density | Training population size | Training population depth | Training population age |
|---|---|---|---|---|---|
| Trait[†] | 4 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| Method[‡] | 1 | 0.22 | 0.02 | 0.21 | 0.06 |
| Design[§] | 2 | <0.0001 | <0.0001 | <0.0001 | <0.01 |
| Trait × method | 4 | 0.03 | 0.14 | 0.56 | 0.31 |
| Trait × design | 8 | <0.01 | <0.01 | <0.0001 | <0.01 |
| Method × design | 2 | 0.26 | 0.11 | 0.64 | 0.52 |
| Error | 8 | | | | |
| Total | 29 | | | | |

[†]Trait is the five traits (beta-glucan, days to heading, groat percent, plant height, or yield).

[‡]Method is the two genomic selection models (ridge regression–best linear unbiased prediction [RR-BLUP] or BayesCπ).

[§]Design refers to different factors for each column of the table. Marker density, number of markers (300, 600 or 900); training population size (100, 200 or 300 lines); training population depth (selection of increasing numbers of lines back in time from the periods 1994–2003, 1998–2003, or 2001–2003); training population age (selection of training populations of equal size from periods of increasing age 1994–1998, 1998–2000, and 2001–2003).

BayesCπ or RR-BLUP, the levels of *design* depend on the design factor being analyzed (i.e., training population size, number of markers, year grouping, or cluster-based grouping), *trait* × *method*, *trait* × *design*, and *method* × *design* are the main effect interaction terms, and the *trait* × *method* × *design* interaction was considered the error term. We recognize that the ANOVA assumption of independence of errors is violated and thus *p*-values are not exact under the null hypothesis. The purpose of this ANOVA is not to test specific null hypotheses but simply to help quantify the relative magnitudes of the factors affecting accuracy.

## Results

### Randomly Selected Training Populations

In all cases, the factor with the strongest effect on accuracy was the trait being predicted (Table 1). Furthermore, this factor interacted in every case with aspects of training population design. In contrast to trait, the two methods we assessed had an impact only on the accuracies of training size but it never interacted with trait or training population design (Table 1).

In general, increasing the number of markers had a positive effect on prediction accuracy (Fig. 1). Maximum accuracy was obtained at the highest density except for groat percentage. The highest increase in accuracy from 300 to 600 and from 600 to 900 markers were both obtained in yield using BayesCπ method with 0.05 and 0.03 increments, respectively. Analysis of variance suggested that not all traits responded equally to an increase in marker density, leading to an interaction between traits and marker density. In particular, groat percentage reached a plateau in accuracy at 600 markers, while for the other traits accuracy continued to increase to the maximum of 900 markers (Fig. 1).

For the standard deviations of accuracies computed from 100 random samples of the training population (data not shown), the values ranged across traits and marker densities between 0.06 to 0.08 for both RR-BLUP and BayesCπ.

Increasing the size of the training population also improved prediction accuracy (Fig. 2). There were differences among the accuracies between traits (Table 1), with β-glucan as the trait with the highest accuracy and yield as the lowest. The accuracies across the three training sizes and traits ranged from 0.23 to 0.49 for BayesCπ and 0.16 to 0.49 for RR-BLUP. There was a steeper increase in accuracy when training population size increased from 100 to 200 than from 200 to 300 lines for all traits except yield (Fig. 2). For instance, β-glucan gained 0.11 (BayesCπ) and 0.09 (RR-BLUP) from 100 to 200 lines, while there was only a 0.05 (BayesCπ) and 0.04 (RR-BLUP) increase in accuracy from 200 to 300 lines. For yield, the increase in accuracy was 0.01 (BayesCπ) and 0.05 (RR-BLUP) between 100 and 200 lines while it was 0.03 (BayesCπ) and 0.05 (RR-BLUP) when the training population was increased from 200 to 300 lines.

The standard deviations produced by BayesCπ were higher (0.08–0.10) across traits than RR-BLUP (0.04–0.06) when the training population size was 100, but were both within 0.04 to 0.08 across methods when the training population included 200 or 300 lines (data not shown).

### Training Populations Constructed from Previous Generations

In practice, training sets will be comprised of previously developed breeding lines. To mimic this approach, the lines were divided based on their first year of entry in the uniform trials and grouped to obtain training population sizes of 90, 180, and 270 lines. Comparison of these training populations will indicate whether it is valuable to include older generations to increase the training population size. The ANOVA for this design (Test-Year in Table 1) indicated that there were differences among the accuracies from different training population sizes grouped according to year. Furthermore, there was also a trait × design interaction, caused primarily by the fact that some traits responded more to increased training population size than did others. The largest gain in accuracy was obtained for β-glucan, in which there was a gain of 0.17 (BayesCπ) and 0.19 (RR-BLUP) when

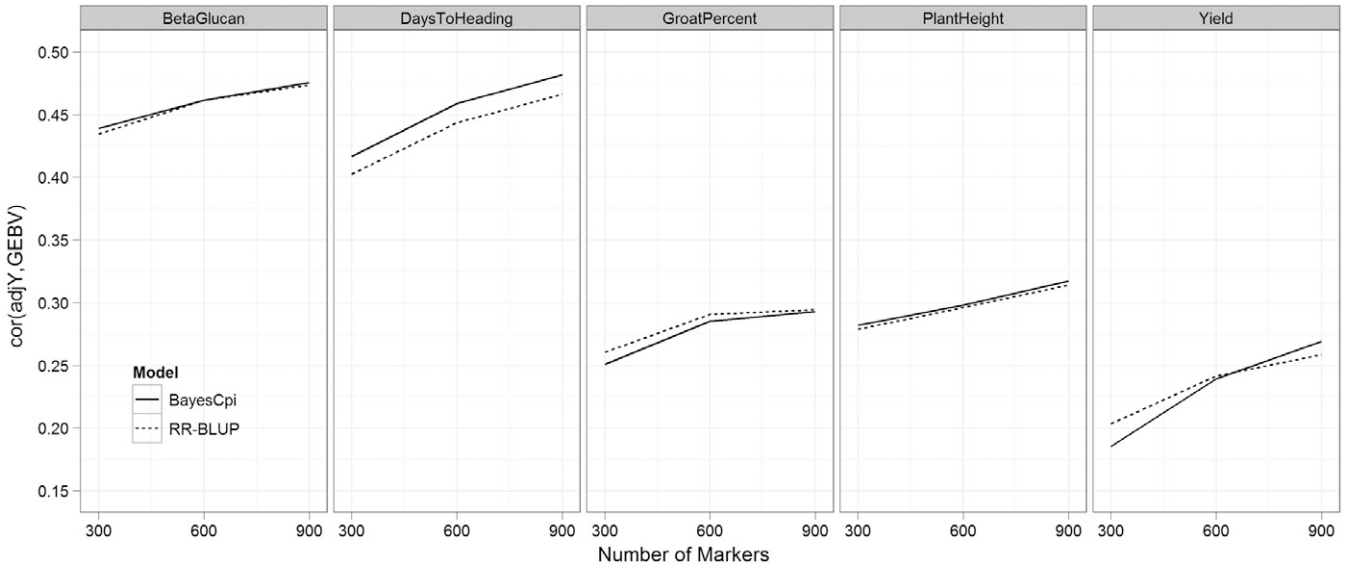


Figure 1. Average accuracies of two genomic selection methods for five traits computed from 100 replicates of randomly selected sets of 300, 600, and 900 markers (*x* axis) included in the model and 300 randomly selected lines used as the training population. The *y* axis is the correlation of population-structure adjusted phenotypic values and the genomic estimated breeding values (GEBV). All correlations shown were significant ($p < 0.05$). RR-BLUP, ridge regression–best linear unbiased prediction.

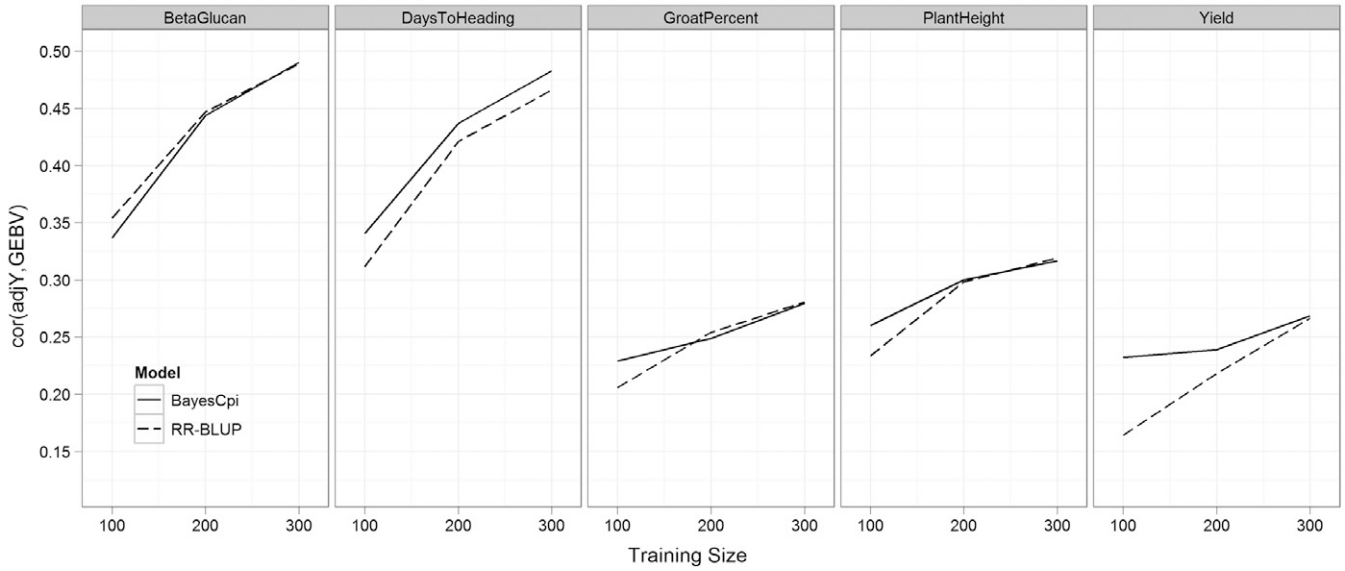


Figure 2. Average accuracies of two genomic selection methods for five traits computed from 100 replicates of randomly selected sets of 100, 200, and 300 lines as training populations (*x* axis) with all 1005 markers included in the model. The *y* axis is the correlation of population-structure adjusted phenotypic values and the genomic estimated breeding values (GEBV). All correlations shown were significant ($p < 0.05$). RR-BLUP, ridge regression–best linear unbiased prediction.

the 1998 through 2003 training population was used instead of the 2001 through 2003 training population (Fig. 3). The lowest gain in accuracy was observed for groat percent, in which there was minimal change in accuracy even when the 1994 through 2003 yr grouping was used as the training population. We also found that using 1998 through 2003 as the training population produced a lower accuracy compared to when 2001 through 2003 was used as the training population for yield. This decrease in accuracy, however, was the only unequivocal decrease resulting from the addition of older phenotypic data to the training population. In other cases, accuracy was constant or increased.

To avoid confounding the effects of training population size and age of training population on prediction accuracy, 90 lines from 1994 through 1998, 1998 through 2000, and 2001 through 2003 were used as the training population. Results showed that most of the statistically not significant accuracies ($p > 0.05$) came from 1994 through 1998 training population. In addition, for this comparison there was also a large design × trait interaction (Table 1). The interaction came from two traits, days to heading and groat percent,

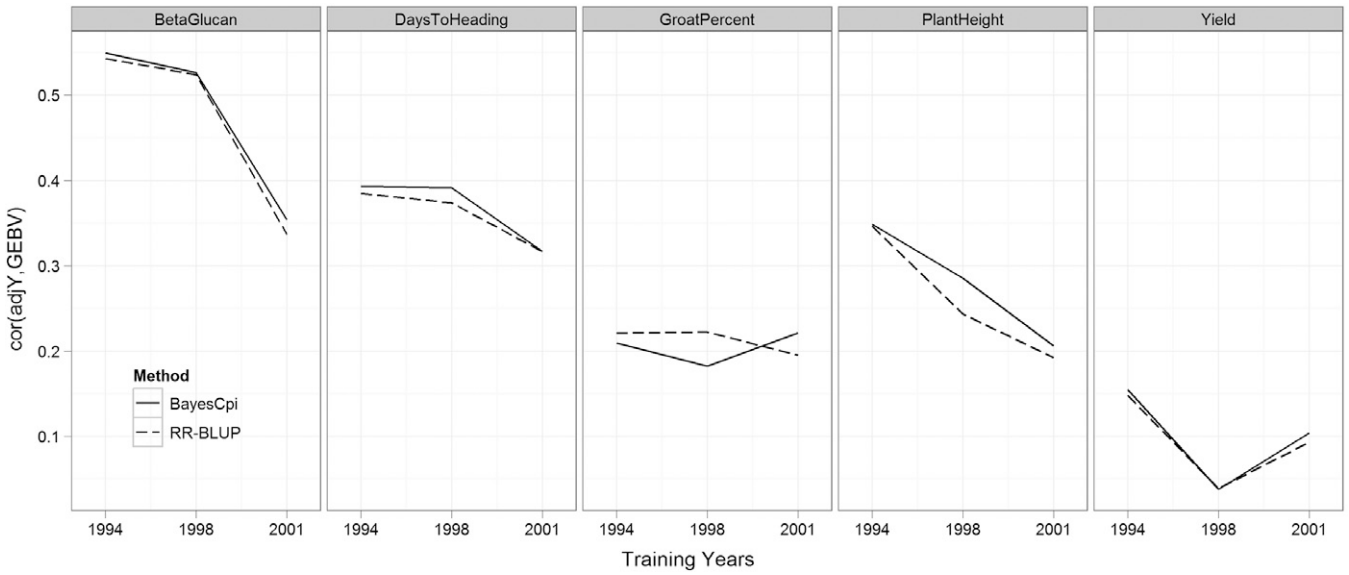


Figure 3. Accuracies for five traits and two genomic selection methods when lines developed during three time periods (1994–2003, 1998–2003, and 2001–2003) were used as the training population to predict lines from 2004 through 2007. The *x* axis shows only the beginning year of each period. The *y* axis is the correlation of population-structure adjusted phenotypic values and the genomic estimated breeding values (GEBV). The minimum correlation that is significant ($p < 0.05$) is 0.14. RR-BLUP, ridge regression–best linear unbiased prediction.

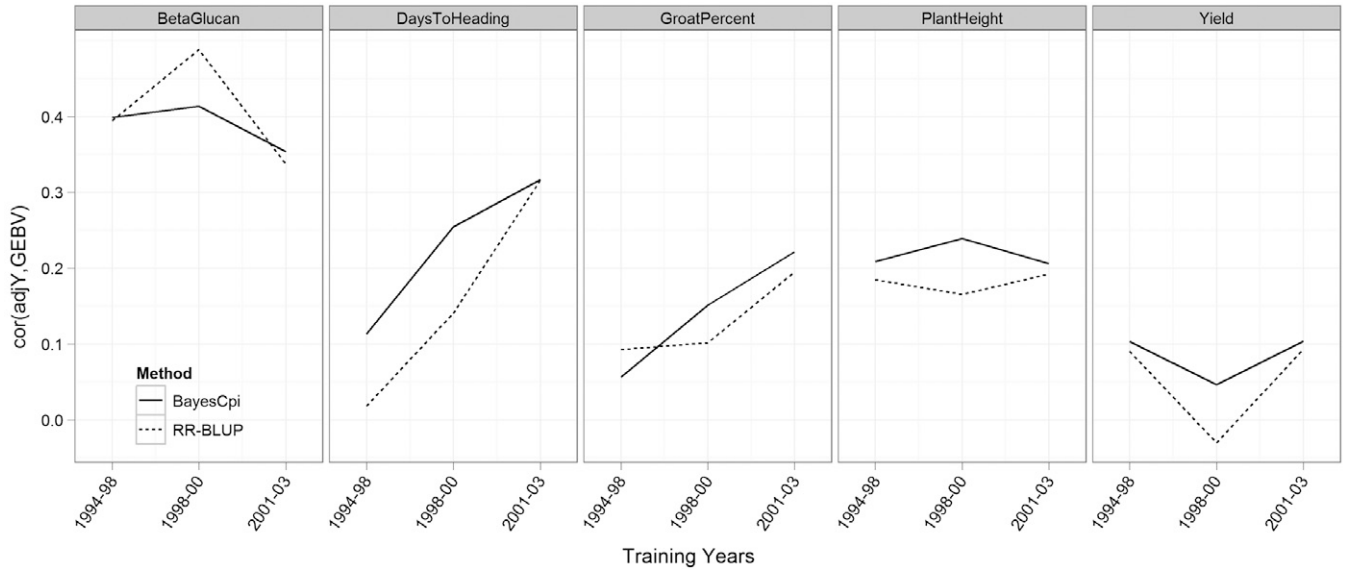


Figure 4. Accuracies for five traits and two genomic selection methods when training populations composed of 90 lines developed during three time periods (1994–1998, 1998–2000, and 2001–2003; *x* axis) were used to predict lines from 2004 through 2007. The *y* axis is the correlation of population-structure adjusted phenotypic values and the genomic estimated breeding values (GEBV). The minimum correlation that is significant ($p < 0.05$) is 0.14. RR-BLUP, ridge regression–best linear unbiased prediction.

for which older training populations led to lower accuracies than recent training populations while for the three other traits, older and recent training populations led to similar accuracies (Fig. 4).

## Training Populations Constructed from Different Subpopulations

To examine the effect of germplasm groupings on the accuracy of GEBV, clusters were used as the training population with a random set of 120 lines from each cluster while the remaining clusters were used as the validation population. Two clusters were also combined each time to form training population sets of 120 and 240 lines. Since C2 and C3 (C23) were more related to each other, they were treated as the *related training population* while C1 and C2 (C12) or C1 and C3 (C13) were treated as the *mixed training population*. Accuracies for single cluster training populations and their combinations are presented in Fig. 5 in which each column of panels corresponds to the validation population. Most

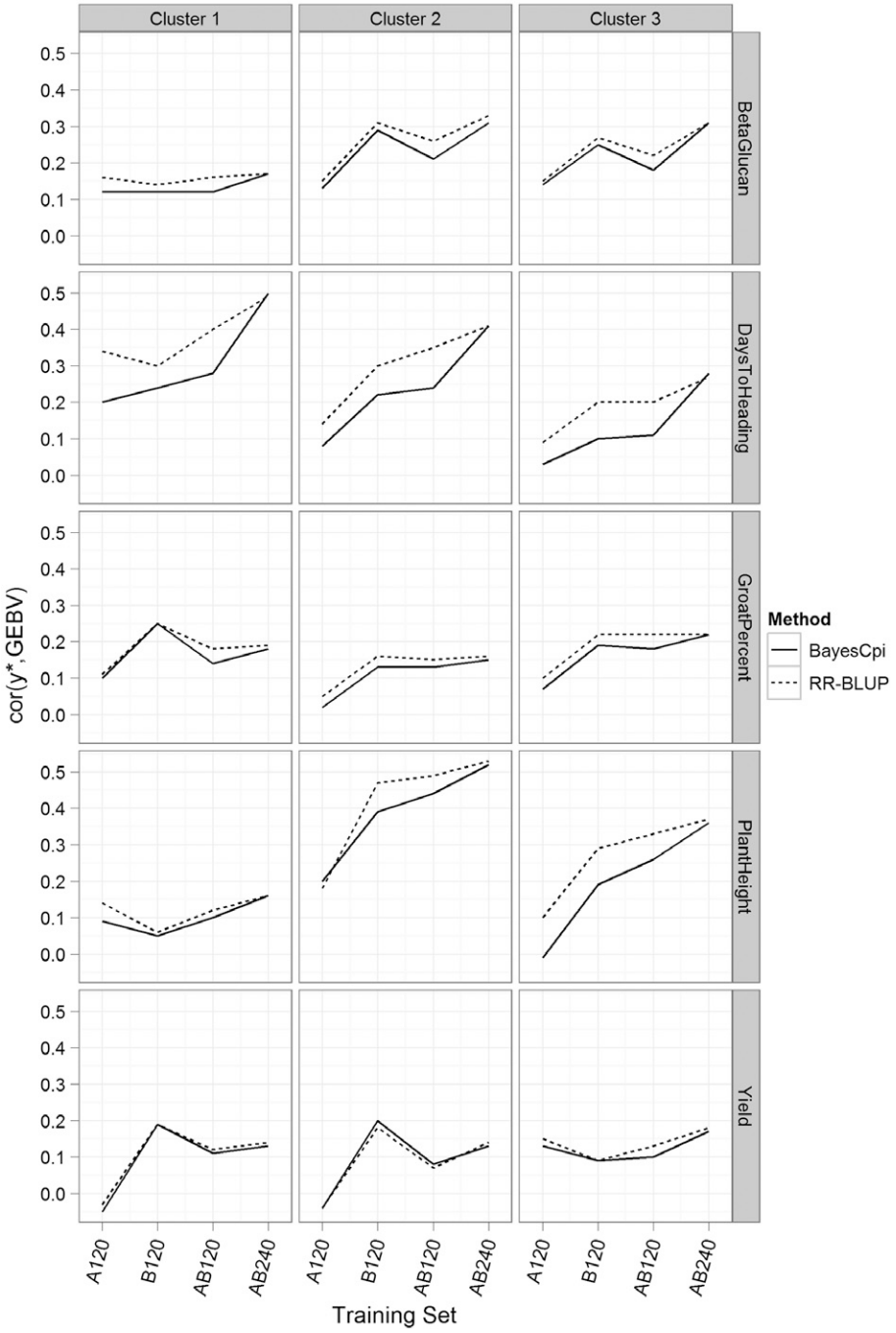


Figure 5. The accuracies of different training populations (*x* axis) across traits (row panels) and validation populations (column panels). *X* axis notation: The letter denotes the cluster from which lines were sampled for the training population, with A for the lower- and B for the higher-numbered cluster (e.g., for C2 as the validation population, A = C1, B = C3, and AB means equal representation of the two clusters). The number gives the training population size. The *y* axis is the correlation of phenotypic values and the genomic estimated breeding values (GEBV). The minimum correlations that are significant ($p < 0.05$) are 0.15, 0.13, and 0.16 for validation populations C1, C2, and C3, respectively. RR-BLUP, ridge regression–best linear unbiased prediction.

of the statistically not significant correlations ($p > 0.05$) were observed when the validation population was C1, followed by C3 and then by C2 (Fig. 5). In this case, the ANOVA showed differences between GS methods and that the method interacted with trait (Table 2). This interaction arose because RR-BLUP was superior to BayesCπ for days to heading across all validation populations and for plant height for the C2 and C3 validation populations, but the two methods performed similarly in all other cases.

Regarding the training population design, we were most interested to determine if related training populations outperformed unrelated training populations and how mixed training populations compared to single-cluster training populations. Because there were trait × design interactions (Table 1), these questions will need

to be addressed trait by trait. Cluster 2 and C3 were more closely related to each other than either was to C1. We therefore expected better prediction when C2 or C3 served as the training population to predict the other than when C1 served to predict C2 or C3. Despite the trait × design interaction, this pattern is constant for every trait (rightmost two columns in Fig. 5: accuracy for B120 is higher than accuracy for A120). In contrast, when C1 was the validation population, there was no reason that either C2 or C3 should generate more accurate predictions and there were generally only small differences between their accuracies across all traits (leftmost column in Fig. 5). We noted also that the highest accuracy in every trait for all 120-sized training populations involved C3 as either a single cluster or part of the mixed training sets (row-wise in Fig. 5). Specifically, the C3 training population had the highest accuracy in β-glucan, groat percent, and yield. In addition, the C23 and C13 training populations had the highest accuracy for days to heading and plant height, respectively.

With respect to the question of *mixed* training populations, the main issue is whether such a training population could generate more accurate predictions than that of the more accurate *pure* training population. The answer to this question varied by validation population and by trait, though overall it resulted in less accurate predictions. Nevertheless, this phenomenon occurred for days to heading for all validation populations and for plant height for the C2 and C3 validation populations (Fig. 5). If *mixed* means also a bigger training population (as would happen if the breeder already had data from two subpopulations and combined them, as represented by the AB240 populations), then accuracies were generally higher than (or at least equal to) the most accurate *pure* training population. This improved accuracy occurred in every case except groat percentage for the C1 and yield for the C2 and C3 validation populations. In general, there was higher gain of accuracy for the BayesCπ method than for RR-BLUP when the training population size was increased from 120 to 240 lines (Fig. 5).

**Table 2. Analysis of variance (ANOVA) *p*-values for factors affecting the accuracies for different validation populations generated from three clusters of oat lines denoted by C3, C2, and C1.**

| Source of variation | df | C3 validation population | C2 validation population | C1 validation population |
|---|---|---|---|---|
| Trait† | 4 | <0.0001 | <0.0001 | <0.0001 |
| Method‡ | 1 | <0.0001 | <0.01 | <0.0001 |
| Design§ | 2 | <0.0001 | <0.0001 | <0.0001 |
| Trait × method | 4 | <0.01 | 0.03 | <0.001 |
| Trait × design | 8 | <0.0001 | <0.001 | <0.0001 |
| Method × design | 2 | 0.79 | 0.32 | 0.04 |
| Error | 8 | | | |
| Total | 29 | | | |

†Trait is the five traits (beta-glucan, days to heading, groat percent, plant height, and yield).

‡Method is the two genomic selection models (ridge regression–best linear unbiased prediction [RR-BLUP] and BayesCπ).

§Design in this table refers to three training populations of 120 lines sampled from clusters other than the corresponding to validation population. For example, design levels for C3 validation population were the training populations C1, C2, and C12 at 120 lines.

## Discussion

This study applied GS methods to empirical data gathered from long-term (1994–2007) multienvironment yield trials for oats in the United States and Canada. The impacts of marker density, training population size, and two GS methods on accuracy of GS were explored. Additionally, the effect of the age of the lines used in the training population and influence of population structure were investigated. Results of this study are encouraging regarding the use of GS in applied breeding programs even with the modest marker density of one marker for every 2 cM on average (1005 markers on a 1890-cM oat map; Wight et al., 2003). While accuracies that we found ranging from 0.27 to 0.50 for training populations of 300 individuals were fairly low and might be insufficient for selection of lines as parents without any further phenotypic information, there are several reasons to believe that accuracies would be higher within breeding programs. First, oat lines in the UOPN are evaluated over a very broad range of environments, including environments outside of the target for which they were bred. Thus, for example yield as measured in this study might be better understood as *broad adaptation yield*. There will be less genetic variance for this broad adaptation yield than for the more narrow adaptation yield that most breeding programs target. Second, the phenotypic data came from highly unbalanced evaluations resulting in more error in the phenotypic observations. This error biases downward the estimated accuracy (Dekkers, 2007; Lorenz et al., 2011). Third, estimated accuracy would have been higher if we had left the effects of structure in the prediction models. The reason for removing those effects is that we were more interested in performance relative to other lines in the same subpopulation than relative to lines in different subpopulations. Finally, we view the largest training population size that we used (300) as a still relatively modest training population.

Accuracy increased with increasing marker density. For β-glucan, days to heading, plant height, and yield, no plateau was reached indicating that more markers would be useful. For groat percentage, however, very minimal increase in accuracy was observed between 600 and 900 markers. It is unclear, however, why a plateau would be reached for some traits but not others. Diversity Array Technology markers may cluster in the oat genome (Tinker et al., 2009). If such clusters happen to coincide with QTL affecting a trait, then a lower marker number would be sufficient to tag all QTL for that trait. Perhaps such an effect occurred with groat percentage. The lower accuracy that was detected for lower marker densities than with higher densities may be explained by the smaller probability of LD between the markers and the QTL when there are fewer markers; hence, only a smaller fraction of genetic variation can be explained (Solberg et al., 2008). Using the

'Kanota' × 'Ogle' comprehensive oat map size of 1890 cM (Wight et al., 2003), this data would indicate that on average there is one marker for every 7 cM when 300 markers are used. This assumes even distribution of markers across the genome, while there was one marker for every 2 cM when all the 1005 markers were used. Simulation (Calus et al., 2008) and empirical (Habier et al., 2010) studies have achieved high GS accuracies using data where average LD between adjacent markers (measured as $r^2$) was 0.20. Newell et al. (2010) explored genome-wide LD in oats and showed that to attain values of $r^2 = 0.20$ between markers, one marker per centiMorgan was needed. These results indicate that we should still see improvements in accuracy up to at least 2000 markers.

There was increasing mean accuracy and lower standard deviations of accuracies with an increase in training population size. This implies that more lines are needed to improve estimates of marker effects and achieve higher accuracies for GS in oats. What is most remarkable about the increase in accuracy with the increase in training population size is that it showed little sign of reaching a plateau for any of the traits analyzed. We hypothesize that this arises from the high level of diversity for the population that we used (Fig. 2). In any event, the result suggests that for training populations that cover several breeding programs, quite large populations will be valuable.

Meuwissen (2009) suggested that an increase in marker density should be coupled with higher training population size to result in higher accuracies. Given the available marker densities in this study, it is more important in the short term to increase the training population size rather than to increase marker density to increase GEBV accuracy.

## Prediction Using Previous Generations as Training Populations

Making training populations based on their chronological entry on the uniform tests can mimic cultivar development processes, in which previous knowledge of the performance of lines can be used to predict future populations. In this kind of design, both LD and the genetic relationships between training population and selection candidates will contribute to accuracy. But since older generations could have a decreasing genetic relationship to recent generations (for this study see Supplemental Fig. S2), the persistence of LD across generations will become more important to maintain accuracy (Habier et al., 2007). The importance of a larger training population size was again emphasized in this design. For all traits that we examined, increasing the training population by adding older lines caused accuracy to either increase or at least remain constant (Fig. 3). The sole exception was yield for the period of 1998 through 2003, though, when, adding even older lines, accuracy again increased. This observation of increased accuracy could be explained by the fact that even quite old lines (e.g., ones from 1994–1998) retained information to predict performance of recent lines (from 2004–2007 in Fig. 4).

We compared equally sized training populations that differed in age and therefore in the time interval between the training and validation populations (Fig. 4). We expected that older training populations would lead to less accurate predictions. In simulation studies (Habier et al., 2007; Zhong et al., 2009) and in a study of Holstein bulls (Moser et al., 2009), when the training and validation populations were several generations removed, accuracy declined. This expectation only occurred for days to heading and groat percentage. Although oat is capable of going through three generations per year, there is a much slower effective generation time in oat breeding programs in which older inbreds may continue to be used as parents for a number of years. If breeding cycle time decreases in the future, through the use of early selection based on genomic prediction, we would no longer expect that such old training populations would retain as much relevant information.

## Prediction of Genomic Estimated Breeding Values in Subpopulations

Most breeding programs have unique groupings of parents that are continuously adapted to produce better populations such as heterotic groups in hybrid breeding or different market classes across a number of crops (e.g., feed versus malt barleys). In this study, groupings in the population were determined by cluster analysis. Cluster 1 was composed mainly of oat lines from Canadian oat breeding programs while C2 and C3 were mostly from the United States. Cluster analysis revealed that C1 was less related to C2 or C3.

As discussed above, the degree of relationship between the training and validation populations affects accuracy of GS (Habier et al., 2007, 2010; Hayes et al., 2009). This effect occurs whether divergence between training and validation populations arises from generations of descent or from population structure. Thus, for the most part, the C2 and C3 clusters predicted each other better than C1 predicted either one (Fig. 5). These findings are similar to that reported by Hayes et al. (2009) for Jersey and Holstein breeds of cattle. This effect of degree of relationship on accuracy was also found within empirical data from four traits of German Holstein Friesian bulls (Habier et al., 2010).

We also found that mixing clusters can offer an alternative design for the training population. When less-related clusters were combined into training populations (i.e., C12 or C13) with the same size as the single clusters, the accuracy was better than the average accuracies for the two single clusters (e.g., average of C1 and C2 versus C12). Using a mixed-subpopulation or multibreed training population has been explored in cattle by Hayes et al. (2009). Their study revealed that multibreed training populations (i.e., Jersey and Holstein) predicted purebred individuals (Jersey or Holstein) with comparable accuracies to the within breed prediction. In the simulation study conducted by de Roos et al. (2009) on training sets composed of two subpopulations (populations A and B), they showed that accuracy of prediction for selection candidates in A was higher if A and B were less divergent than when A and

B were highly divergent. The empirical study of Daetwyler et al. (2010a) in sheep demonstrated that the breed of the selection candidates that was most represented in multibreed training populations achieved higher accuracies. Similar to what was found in this study, C12 or C13 training populations provided higher accuracy than C23 on average, because the former had related lines between the training and validation populations while the latter had training and validation populations that were less related.

Accuracy can be increased with higher marker density even if training sets and selection candidates are highly divergent (de Roos et al., 2009). Meuwissen (2009) also suggested that in predicting unrelated individuals, a substantially larger training data set and a higher marker density are required to obtain high accuracies. These results lead to the recommendation that a single large mixed training population with a higher marker density would offer a better solution than multiple training populations, each serving one germplasm group. Higher marker density will help to increase the probability of finding markers that are in consistent LD with the same QTL across the different subpopulations (Daetwyler et al., 2010a). The focus of this strategy will be GS model building in which consistent historical LD across subpopulations is explored rather than just within-subpopulation LD.

We hypothesized that doubling the training set size would be less beneficial when the training population was composed of related individuals (e.g., C23) than when it was composed of unrelated individuals (e.g., C13). That effect was observed for β glucan, plant height, and yield but not for days to heading and groat percent (data not shown). Results for increasing marker densities were likewise inconclusive. We believe a larger total experiment size would be needed to detect these effects.

## Global Comparison of BayesCπ and Ridge Regression–Best Linear Unbiased Prediction for all Training Designs

Training population designs used in this study found that neither GS method was consistently better in terms of accuracy. Simulation studies of Jannink (2010) showed that the difference of these two methods in terms of genetic gain were very small under low (0.20) and medium (0.50) heritabilities and varying training population size of 200 or 1000. However, in this study BayesCπ was consistently better or the same than RR-BLUP for days to heading across different marker density and randomly versus chronologically selected training populations (Fig. 1 through 4). It was also observed that for small training set sizes (90–100 lines in our case), BayesCπ outperformed RR-BLUP in four out of five cases for randomly selected training sets (Fig. 2) and in 13 out of 15 cases for chronologically selected training sets (Fig. 4). Similar results under small training population size were obtained by Meuwissen (2009) though conflicting observations on the performance of these types of models with small training sets have also been reported (Daetwyler et al., 2010a; Habier et al., 2010).

Hayes et al. (2009) conceptualized the performance of multisubpopulation training populations as dependent on the detection of ancestral LD that is common across subpopulations. This idea would suggest that methods that capture marker-QTL LD will be more effective than methods that model genetic relationships between the training and validation populations (see Habier et al., 2007, and Zhong et al., 2009, for a discussion of these two components of GS accuracy). Thus, we expected BayesCπ to outperform RR-BLUP in analyses where the training population came from a different subpopulation than the validation population or where the training population was mixed. In fact, we observed the opposite: RR-BLUP was better than BayesCπ in the cluster-based design for a training population comprised of 120 lines. We have no compelling explanation for this observation though we note that, in these cross-subpopulation analyses, we could not include a term to account for population structure in the genomic prediction linear model. Failure of line clustering to account for all effects of subpopulation structure may therefore have played a role.

The difference in terms of average accuracy and standard deviations between BayesCπ and RR-BLUP decreased in larger training populations across different designs in this study. This was similar to the result of Meuwissen (2009) in which BayesB (related to BayesCπ) had similar accuracy with GBLUP (equivalent to RR-BLUP) when using larger training populations. These two methods differ in their assumptions of variance of marker effects; the former uses unequal variance for each marker while the latter assumes that all markers have equal variance. At constant heritability, RR-BLUP is insensitive to genetic architecture (i.e., the number of QTL and the distribution of their effects), while the accuracy of Bayesian methods improves as the number of QTL decreases and their effects increase (Luan et al., 2009; Daetwyler et al., 2010b).

## Implications for Plant Improvement Programs

Accuracy as a component of response to selection can be used to predict the future gains using GS. As an example, accelerated breeding for β-glucan, a compound found in oats that has been shown to have positive health benefits (U.S. Food and Drug Administration, 2010), can benefit from GS. Beta-glucan is a polygenic trait governed by genes with mainly additive effects and heritability ranging from 0.27 to 0.58 (Cervantes-Martinez et al., 2001). In a typical phenotypic selection program, β-glucan content is evaluated every year from seeds of replicated plots during the summer season. To adapt a GS strategy for β-glucan improvement, in which there are two cycles of selection that can be done in 1 yr (e.g., Jannink, 2010), an accuracy equal to 1/2 $h$ may be enough to justify GS conducted twice a year. Assuming a heritability of 0.5 (h = 0.71) versus a GS accuracy of $r$ = 0.5, GS will lead to around 40% more gain than phenotypic selection per unit time. Genomic selection, however, should be further validated in breeding programs with several generations to determine both advantages and disadvantages and

modifications that could potentially maximize genetic gain. As mentioned, GS in plant breeding can be applied in broad-based populations such as this study and Heffner et al. (2011) or in narrow-based populations such as biparental populations (Lorenzana and Bernardo, 2009). Applications of GS with respect to these types of populations differ because of the extent LD: marker density requirements for biparental populations are much lower than for a set of lines with broad genetic diversity. Furthermore, population structure is of no concern in biparental populations since all individuals are equally related. Finally, the time requirement of GS model building will be greater in biparental populations due to the fact that every biparental population will need phenotypic data before model training (Heffner et al., 2011). Specific studies will need to be implemented to determine which GS process is best suited for the crop of interest.

# Appendix

## Appendix 1. Pre–Genomic Selection Analysis of Phenotypic Data in SAS.

Mixed Model: $y = \mathbf{X}b + \mathbf{Z}u + e$,

in which $y$ is the phenotypic data from unbalanced multienvironment trials, $\mathbf{X}$ is the design matrix for environments, $b$ is the fixed environmental effects, $\mathbf{Z}$ is the design matrix for oat lines, and $u$ is the random oatline effects.

For GS purpose in this study, $y^* = u$ + overall mean was treated as the observed value of each oatline.

## Supplemental Information Available

Supplemental material is available free of charge at http://www.crops.org/publications/tpg.


### Acknowledgments

This research was funded by the United States Department of Agriculture, National Institute of Food and Agriculture, grant 2008-55301-18746. We thank Adrienne Moran Lauter for laboratory work and Nick Tinker and Charlene Wight for providing oat pedigree data.


### References


Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos, and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561. doi:10.1534/genetics.107.080838

Cattell, R.B. 1966. The scree test for the number of factors. Multivariate Behav. Res. 1:245–276. doi:10.1207/s15327906mbr0102_10

Cervantes-Martinez, C.T., K.J. Frey, P.J. White, D.M. Wesenberg, and J.B. Holland. 2001. Selection for greater β-glucan content in oat grain. Crop Sci. 41:1085–1091. doi:10.2135/cropsci2001.4141085x

Chernyshova, A.A., P.J. White, M.P. Scott, and J.-L. Jannink. 2007. Selection for nutritional function and agronomic performance in oat. Crop Sci. 47:2330–2339. doi:10.2135/cropsci2006.12.0759

Colleoni-Sirghie, M., J.-L. Jannink, I.V. Kovalenko, J.L. Briggs, and P.J. White. 2004. Prediction of β-glucan concentration based on viscosity evaluations of raw oat flours from high-β-glucan and traditional oat lines. Cereal Chem. 81:434–443. doi:10.1094/CCHEM.2004.81.4.434

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724.

Daetwyler, H.D., J.M. Hickey, J.M. Henshall, S. Dominik, B. Gredler, J.H.J. van der Werf, and B.J. Hayes. 2010a. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Anim. Prod. Sci. 50:1004–1010. doi:10.1071/AN10096

Daetwyler, H.D., R. Pong-Wong, B. Villanueva, and J.A. Woolliams. 2010b. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031. doi:10.1534/genetics.110.116855

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J.M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigrees. Genetics 182:375–385. doi:10.1534/genetics.109.101501

de Roos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of genomic breeding values across multiple populations. Genetics 183:1545–1553. doi:10.1534/genetics.109.104935

Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x

Diversity Arrays Technology. 2011. Plant DNA extraction protocol for DArT. Available at http://www.diversityarrays.com/sites/default/files/pub/DArT_DNA_isolation.pdf (verified 26 May 2011). Diversity Arrays Technology, Pty Ltd., Yarralumla, Australia.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman Technical and Scientific, Essex, UK.

Garrick, D.J., J.F. Taylor, and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55

Habier, D., R. Fernando, and J. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397.

Habier, D., J. Tetens, F.R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5. doi:10.1186/1297-9686-42-5

Hamblin M.T., T.J. Close , P.R. Bhat, S. Chao, J.G. Kling, K.J. Abraham, T. Blake, W.S. Brooks, B. Cooper, C.A. GriVey, P.M. Hayes, D.J. Hole, R.D. Horsley, D.E. Obert, K. P. Smith, S.E. Ullrich, G.J. Muehlbauer, and J.-L. Jannink. 2010. Population structure and linkage disequilibrium in U.S. barley germplasm: Implications for association mapping. Crop Sci. 50:556–566.

Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41:51. doi:10.1186/1297-9686-41-51

Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Gen. 4:65–75. doi:10.3835/plantgenome.2010.12.0029

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9–112.

Jannink, J.-L. 2010. Dynamics of long-term genomic selection. Genet. Sel. Evol. 42:35. doi:10.1186/1297-9686-42-35

Jannink, J., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. Brief. Funct. Genomics Proteomics 9:166–177. doi:10.1093/bfgp/elq001

Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723. doi:10.1534/genetics.107.080101

Kizilkaya, K., R.L. Fernando, and D.J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88:544–551. doi:10.2527/jas.2009-2064

Lorenz, A., S. Chao, F. Asoro, E. Heffner, T. Hayashi, H. Iwata, K. Smith, M. Sorrels, and J.-L. Jannink. 2011. Genomic selection in plant

breeding: Knowledge and prospects. p. 77–123. *In* D.L. Sparks (ed.) Advances in agronomy. Academic Press, San Diego, CA.
Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151–161. doi:10.1007/s00122-009-1166-3
Luan, T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T.H.E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126. doi:10.1534/genetics.109.107391
Matthews, D.E. 2011. Quick queries for the UOPN and QUON. Available at http://wheat.pw.usda.gov/GG2/uopnquery.shtml (verified 26 May 2011). GrainGenes, Albany, CA.
Meuwissen, T.H.E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. doi:10.1186/1297-9686-41-35
Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.
Moser, G., B. Tier, R.E. Crump, M.S. Khatkar, and H.W. Raadsma. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet. Sel. Evol. 41:56 doi:10.1186/1297-9686-41-56
Newell, M.A., D. Cook, N.A. Tinker, and J.-L. Jannink. 2010. Population structure and linkage disequilibrium in oat (*Avena sativa* L.): Implications for genome-wide association studies. Theor. Appl. Genet. 10.1007/s00122–010–1474–7.
Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. doi:10.1038/ng1847
R Development Core Team. 2009. R: A language and environment for statistical computing. Available at http://www.r-project.org (verified 17 May 2011). R Foundation for Statistical Comput., Vienna, Austria.
SAS Institute. 2008. SAS/Stat user's guide. SAS Institute, Cary, NC.
Solberg, T.R., A.K. Sonesson, J.A. Woolliams, and T.H.E. Meuwissen. 2008. Genomic selection using different marker types and densities. J. Anim. Sci. 2008(86):2447–2454. doi:10.2527/jas.2007-0010
Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger. 2008. Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754. doi:10.1534/genetics.107.079707
Taylor, J.R. 1997. p. 128–129. An introduction to error analysis: The study of uncertainties in physical measurements. University Science Books, Sausalito, CA.
Tinker, N.A., A. Kilian, H.W. Rines, A. Bjornstad, C.J. Howarth, J.-L. Jannink, J.M. Anderson, B.G. Rossnagel, C.P. Wight, D.D. Stuthman, M.E. Sorrells, G.J. Scoles, P.E. Eckstein, H.W. Ohm, E.W. Jackson, S. Tuvesson, F.L. Kolb, S.J. Molnar, O. Olsson, M.L. Carson, A. Ceplitis, J.M. Bonman, L. Federizzi, and T. Langdon. 2009. New DArT markers for oat provide enhanced map coverage and global germplasm characterization. Biomed. Central (BMC). Genom. 10:1471–2164.
Toosi, A., R.L. Fernando, and J.C.M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32–46. doi:10.2527/jas.2009-1975
U.S. Food and Drug Administration. 2010. Health claims: Soluble fiber from certain foods and risk of coronary heart disease. Available at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=101.81 (verified 19 May 2011). U.S. FDA, Silver Spring, MD.
Wight, C.P., N. Tinker, S.F. Kianian, M.E. Sorrells, L.S. O'Donoughue, D.L. Hoffman, S. Groh, G.J. Scoles, C.D. Li, F.H. Webster, R.L. Phillips, H.W. Rines, S.M. Livingston, K.C. Armstrong, G. Fedak, and S.J. Molnar. 2003. A molecular marker map in 'Kanota' × 'Ogle' hexaploid oat (*Avena* spp.) enhanced by additional markers and a robust framework. Genome 46:28–47. doi:10.1139/g02-099
Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, and J.B. Holland. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702
Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.-L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182:355–364. doi:10.1534/genetics.108.098277